Offering the Option of Don’t Know

Overview

Your company or university may already have a standard for whether to allow a “don’t know” response option or not with image rating questions. Regardless, it is worth a review of some of the literature to date, which is extensive. Our own expectation going into the study was that a don’t know response option *should* be included, but some of the readings as well as some of the results from our research have challenged that assumption.

Prior Work

Two types of errors may occur with don't know (DK). One – a respondent not offered a DK may give a response not reflecting his or her true opinion (or lack thereof). Two – a respondent offered a DK may choose it even when they *do* have an opinion. Of course, there is no guarantee that with or without a DK respondents will give or even be able to give 100% valid and reliable responses.

Feick (1989) in his own review helps us think about where the DK might be coming from – qualities of the respondent vs. qualities of the questionnaire. Older, less educated, nonwhite, lower income categories and women are more likely to use DK. Questions that are: more complex, require the respondent to think far ahead, poorly constructed, or on a topic of little interest or familiarity to the respondent all may increase the incidence of DK. To avoid the nonrandom bias introduced by a DK, one solution he suggests is to eliminate them altogether.

Feick notes other authors have seen a halo effect where respondents answer based on their general feelings rather than specific attitudes, including giving opinions on non-existent entities. Feick went further into DK with latent class analysis but for our paper we just want to extract elements of his lit review.

Opposition to Don't Know

Not every author agrees that DK should be included. This is part of what makes it interesting. Krosnick and a host of distinguished co-authors (2002) ran nine experiments in three household surveys to test respondents' use of DK, questioning if adding a DK would only draw those who were otherwise giving meaningless data or if it would also entice those who might have an opinion and would have otherwise given it. They note:

If offering a no-opinion option reduces non-attitude reporting, it should strengthen correlations between opinion reports and other variables that should, in principle, be correlated with them. If non-attitude reports are random responses, then offering a no-opinion option should reduce the amount of random variance in the attitude reports obtained.

Krosnick's theory of satisficing suggests respondents may be unmotivated to answer particular questions, especially complex ones, and so may choose "don't know" simply as a way to continue the interview, especially when cued that this option exists. Krosnick hypothesized that omitting the no-opinion option would cause the strong "satisficers" to give their substantive answer instead and eliminate this shortcut to cognitive laziness, as it were.

Nine studies later, they concluded that including a no-opinion option did not increase the quality of the data, but instead that the respondents drawn to no-opinion options "would have provided substantive answers of the same reliability and validity as were provided by people not attracted to those options." One suggestion they make to researchers who still wish to include a no-opinion response is to probe respondents who say DK with whether they lean one direction or another. This would reduce the satisficing (if it is happening) in encouraging the respondent to think and not allowing an easy out.

In favor of including Don't Know

In early work, Converse (1970) suggests that respondents will give random responses if they don't know but don't want to appear ignorant.

Stieger, Reips and Voracek (2007) notes that if we force a respondent to respond, we may induce reactance – and that this is a possible outcome of the (relatively) new mode of online surveys. Reactance is an emotionally triggered state in response to excessive control where the individual feels their freedom is threatened, and therefore attempts to re-establish their freedom by acting in the opposite mode of what the situation requires or requests. Reactance theory was first proposed by Brehm (1966), and is the idea behind the popularized reverse psychology. Stieger et al. hypothesize the lack of a DK will lead to respondents deliberately giving misleading or inaccurate responses, or to simply dropping out of the study altogether.

In Stieger's methodology with 4,409 University of Vienna students, test group respondents who attempted to advance without filling in a question on the infidelity questionnaire would receive an error screen asking them to completely fill in the questionnaire. The event was logged for later analysis. Control group respondents did not receive the error page.

Instructionally, 394 respondents dropped out immediately after receiving their first error page, particularly on the "demographics" page (it appears all demographics were collected on one screen). Another 121 dropped out later. Only 288 received an error page and still completed the questionnaire. The dropout rate of those who did not attempt to skip a page was 18.6% vs. 64.1% of those who did so at least once. In addition, the authors did find indicators of reactance – the data for respondents after receiving an error page was significantly different from the data for those same questions for respondents who did not. In addition, Stieger found men dropping out faster than women in the forced-response condition (this author supposes it might have something to do with the content material and might not be a finding with less provocative questions).

In their discussion, Stieger et al. would like to distinguish between "good dropout" and "bad dropout". If respondents are not going to give us quality data due to poor motivation then we wish them well but don't want to include them in our study. Bad dropouts may be due to inadequate questionnaire design, programming errors, lack of feedback on progress, et cetera. A very low dropout rate may in fact be a bad thing if we're keeping 'bad' respondents in our data.

Finally, Stieger suggests criteria for forced-response design:

1. It is necessary to have a complete set of replies from the participants (e.g., semantic differentials, multivariate analyses, pairwise comparisons, required for skip patterns)
2. A high response rate is expected and so dropout is not a concern
3. The distribution of respondents' sex is not a main factor in the study

Friedman and Amoo (1999) propose that if subjects are undecided and have no 'out' they will probably select a rating from the middle of the scale, biasing the data in two ways: "(a) it will appear that more subjects have opinions than actually do (b) the mean and median will be shifted toward the middle of the scale." They also remark on the usefulness of %

don't know, especially in political polling where the previously undecideds can change an election. Note the probably italicized above, as Friedman did not back this assertion up with data.

In favor of including Don't Know, *delineated*

In an intriguing bit of research on online surveys, Tourangeau, Couper and Conrad (2004) investigate the placement on the screen of "nonsubstantive" response options such as don't know in relation to their substantive counterparts.
Tourangeau et al present results for three different interpretative heuristics they believe respondents are using that may lead to misreadings of survey questions:

1. Middle means typical
2. Left and top mean first (either worst or best)
3. Near means related

Note that prior research (cited by Tourangeau) already supports the first heuristic, and must be taken into account when delivering closed-ended range questions to respondents in lieu of open numeric questions. An implication of "near means related" is higher correlations in items presented as a grid than those presented on separate screens.

The authors ran two surveys testing the middle means typical in 2001 and 2002, through Gallup, with 2,987 interviews of 25,000 invitations in the first study and 1,590 of 30,000 in the second study. Respondents received an attitude question with a vertically presented scale, five substantive points ordered high to low ('far too much' to 'far too little'), followed by both a "don't know" and a "no opinion". Test groups had a short divider line, a long divider line, or a space between the five and the two. The control group saw all seven options contiguously.

In all cases, the means are statistically closer to the "far too little" point when there is no separation between the substantive and non-substantive responses. However, a side effect of setting apart the non-substantive responses in this case led them to being chosen more often.

Tourangeau followed up with an experiment simply adjusting the spacing in the scale question, horizontally, either visually crowding some of the responses to one side or spacing them evenly. Again, the means moved towards the visual center, not the labeled center of the scale.

They conclude,

Our results indicate that [respondents] may also make unintended inferences based on the visual cues offered by the question. Basing their reading on the questions' visual appearance, respondents may miss key verbal distinctions and interpret the questions in ways the survey designers never intended.

Our Study

We had four questions to answer with allowing (or disallowing) a don't know response option:

- Did we intolerably increase the level of noise in the data by removing it?
- Were those who used DK different from those who didn't, thus possibly biasing the data?
- How would our respondents behave if they did not have DK?
- How would our multivariate applications look in the non-DK situation?

Respondents rated three brands on a five-point scale (the one very top firm, world class, stronger than most, average, weak) with or without don't know for sixteen attributes regarding brands in the financial industry. Figure 4 presents the questions used in both parts of this study.

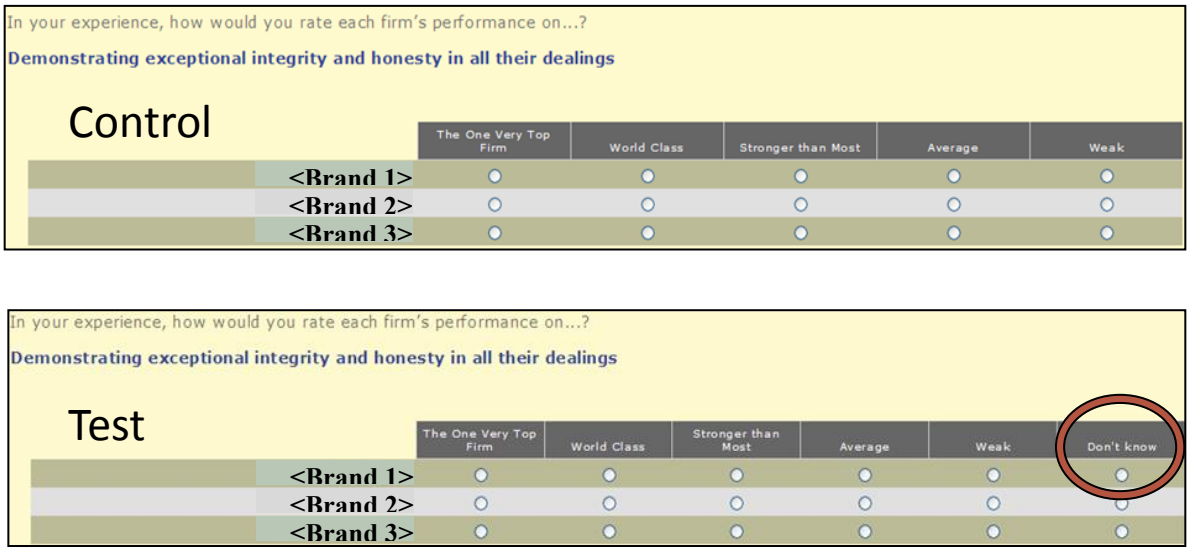

Figure 4:
Control and Test questions for Don't Know Study

To qualify for this study, respondents needed to have voted, have a certain minimum level of investments and income, and be actively involved in expressing their opinions on financial issues. The screener averaged three minutes to complete, followed by a six-minute questionnaire (half of which was the brand rating series). The e-Rewards panel provided the online sample.

Differences in the control and test groups were as follows:

	Control	Test
Don't know	Absent	Present
Interviews (n)	272	155
Field dates	November 5 – December 2, 2008	December 5 – 12, 2008

Samples differed insignificantly on age, income, education, gender.

Results

If our goal had been to avoid frustrating responses, removing DK would have caused us to miss the target. Of respondents not given the option to select DK, over 20 commented on it negatively when asked at the end of the survey to evaluate the questionnaire, for example:

There was no option to say I don't know, forcing me to make choices on some questions I was not qualified to answer.

Just because I indicated I was 'familiar' with some companies doesn't mean that I'm in a position to answer such detailed questions about them. I often felt that 'don't know' or NA should have been an option.

There should always be an opt-out response on questions as the respondent may not have a response and then is forced to respond if there is no opt out response. This is very basic stuff.

When DK was present, over half the test group respondents took advantage of it at least once in the 16-attribute section (and none complained). But what about those who didn't have DK - what did they do?

A Series of Hypotheses

First let's compare actual results in allowing a DK or not (stacked data across all attributes).

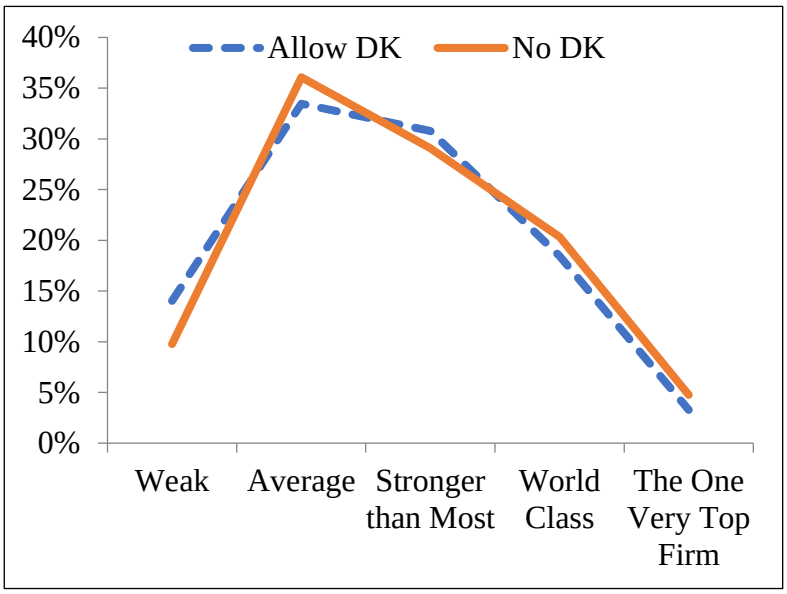

Chart 8

The respondents without a DK response option did select “average’ slightly more often than those who were provided a DK option. To understand the dynamic more deeply, we then used several simulation designs in an attempt to replicate analytically the process that these respondents were going through mentally.

What if they plumped for the most neutral response of “average”, as Friedman and Amoo (1999) suggested they might? As Chart 9A below demonstrates where the DK value has been replaced by “average”, that’s clearly not happening. What if the respondents just randomly chose one of the five responses, as Converse (1970) says might happen when respondents would prefer not to appear ignorant? Chart 9B replaces the DKs with random responses. This simulation is closer to the actual responses, but still varies by a significant degree.

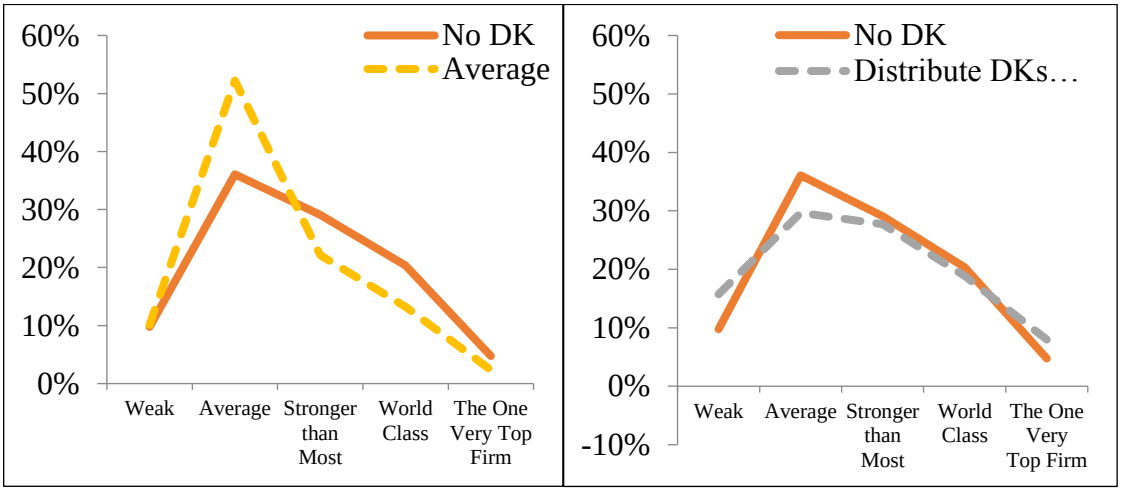

Chart 9A and 9B

Another possibility is the respondent restricts their choices to the middle values. We’re not providing a graph illustrating this as the standard deviations were virtually identical between the DK and No DK group, thus allowing us to discard it as a hypothesis.

Finally, as Feick and other authors suggest, the respondents may infer (or impute) from what they know generally about the brand to score an attribute. Using a simplistic imputation, replacing the DKs with the integer nearest to the mean score *of all other attributes for that respondent*, we obtain the results shown in Chart 10.

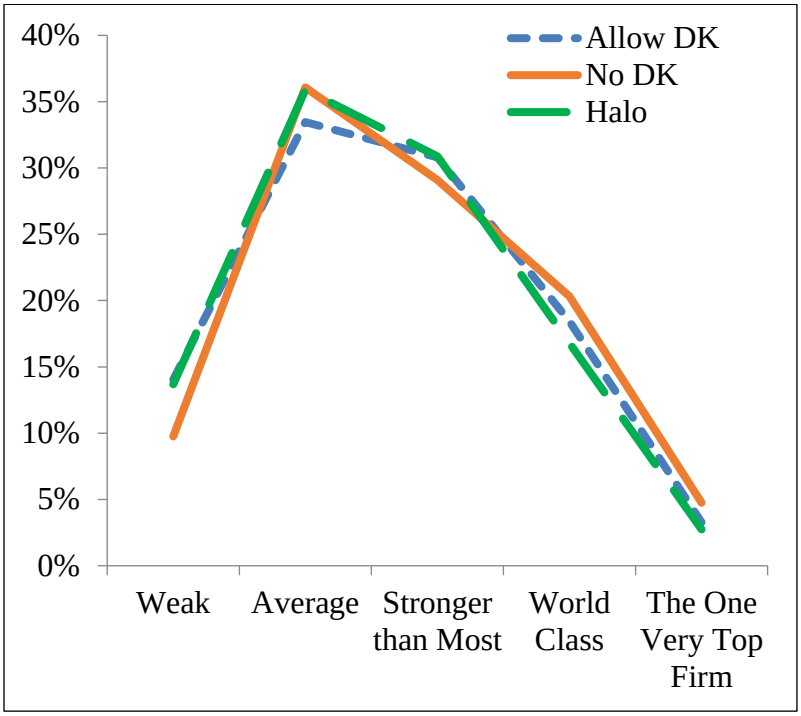

Chart 10

As you can see, populating the DKs from the "Allow DK" group with what they generally knew about the brand lines up very closely with the No DK group. And this was not a small percent; 28% of the observations (brand x attribute) were DK.

These simulations give us more confidence that when DK is not available as a response option, respondents will act in good faith and impute reasonably.

Back to Real Data

We then looked at how respondents with access to the DK response option behaved. Do they also act in good faith?

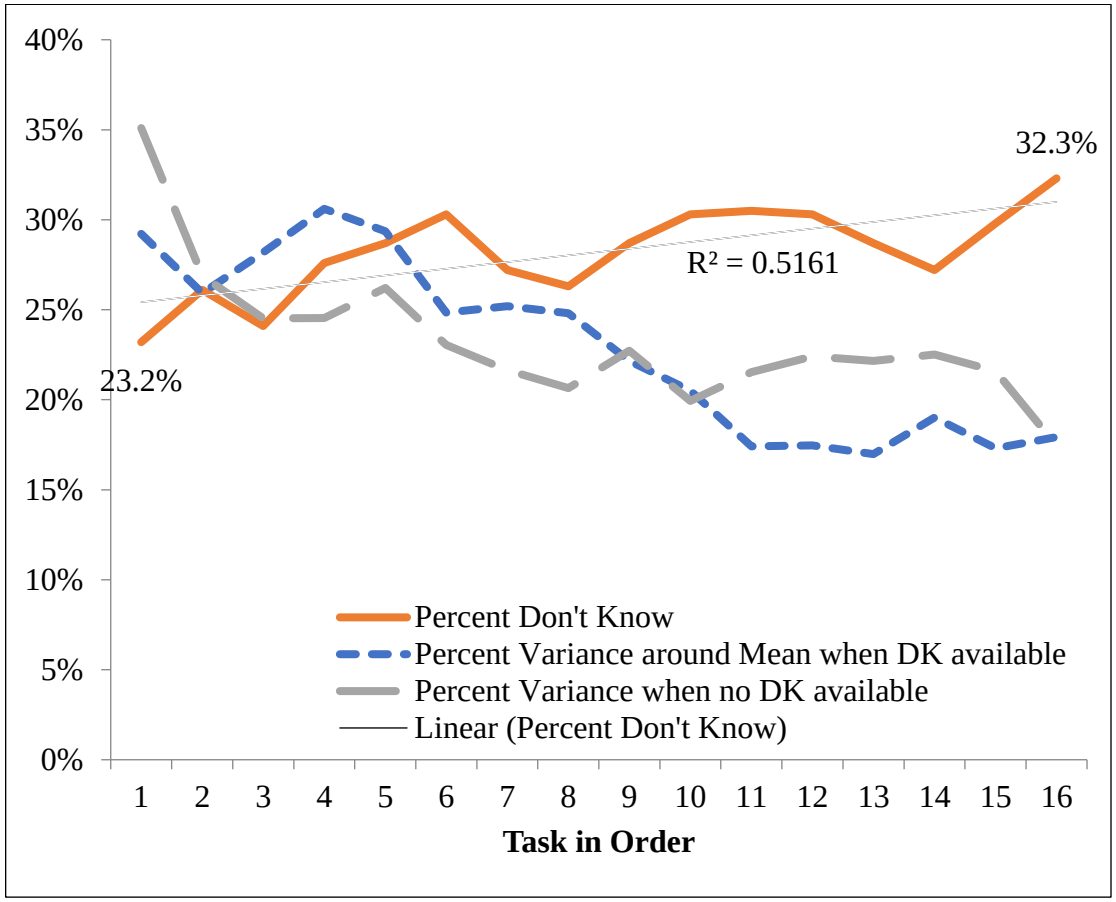

Chart 11

As seen in Chart 11, respondents used don't know significantly more often as the series of grid questions progressed ($p<.001$ for linear contrast), at almost a 40% increase compared to the beginning. And both groups gave answers with less and less variance as time went on. Respondents allowed a DK do not always act in good faith.

In addition, we saw some differences in means, even when controlling for familiarity. The mean value was lower for the DK group than the non-DK group six out of sixteen times when "somewhat familiar" with the brand and four out of sixteen when "very familiar" with the brand.

Two possible explanations come to mind:

- As we know the financial industry did not have a very good 2008, either financially or in the public perception. Our test group was fielded a bit later than the control group and may have had exposure to even more bad news at that point.
- The DK response option was placed contiguous with all other points, with no distinguishing visual features whatsoever. Corresponding to Tourangeau and co-authors' findings (2004), this unfortunate placement may have visually shifted the mean in respondents' minds towards the right, the lower level of the scale. (As an aside, Cambia intentionally uses an unbalanced scale as our clients really only want to be "The one very top firm" but this may throw off panelists those who expect the middle point to be average.)

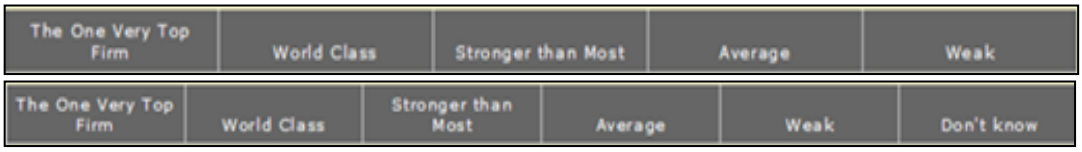

Figure 5: Revisiting the scales

As one of our goals was to compare the impact of DK on the variables that mattered, we correlated the attributes (stacked) with several outcome variables. Note that this methodology was chosen over regression as the multicollinearity made regression comparisons untenable.

	3 brands No DK	3 brands DK	6 brands No DK
How likely are you to recommend this firm to a professional colleague who is looking to do business with a firm in this industry	.620	.601	.621
How likely are you to invest with this firm	.581	.582	.623
Extent to which you want to see firm succeed	.443	.381	.373

Table 1: Correlations of Attributes (as composite) with Outcome Variables

The results showed that allowing respondents a DK response option did not strengthen the correlation with outcome variables; if anything, it decreased the correlation. This was similar to Krosnick's findings. For further comparison, the 6-brand test is also shown, and the results are parallel to the DK set.

To round out the analyses, we looked at the demographics of those who used DK and those who didn't. As mentioned in Feick (1989), women used don't know more often than men (33% to 26%). Other demographic variables did not follow published outcomes for DK:

- Those more highly educated used don't knows more often than those with less education (Krosnick, 2002 had the reverse)
- Higher income respondents used more don't knows
- Age was virtually unrelated to use of don't know (again, Feick had younger using don't know more often)

Note that as a set of (older) panelists with $100K+ in investments our sample is not representative of the general population.

Conclusion

We came into this research effort with an expectation that, in general, providing a don't know response is advantageous over omitting it. However, for this study, for this context, this sample, this time period, this content area… taking away DK

- did not increase noise,
- did not change the distribution across attributes (though some means were affected), and
- did not negatively impact the relationships with outcome variables.

Respondents seemed to do a fine job of self-imputation. Had the variables been less intercorrelated so that we were able to do multivariate analyses with the data, we would have had an easier time with the non-DK group than the DK group, avoiding the need for imputation. Some respondents with the DK response option clearly used it when they did not "need" to, if not at the beginning of the series then certainly by the end.

But, one must weigh the relative benefits of having a clean data set with the drawbacks of potentially frustrating respondents (who also may be prospective clients). The questionnaire was so short that we did not have an appreciable number of partials to contrast if the DK absence were leading to greater dropouts. But our questionnaire feedback question provided them an opportunity to vent, and vent they did.

We also saw some demographic differences in DK usage which could lead to unintentionally biased data.

Our takeaway is to graphically alter where we place the DK so it is clearly visually separated from the grid (and perhaps in a smaller font to de-emphasize it). And should we have a request to omit DK, we can feel marginally more comfortable this will not negatively affect the results.